R E P O R T R E S U M E S

ED 003 882 48

DEVELOPMENT OF A DEVICE TO RECORD GRAPHICALLY INTONATIONS OF SPEECH AS THEY ARE PERCEIVED BY A LISTENER. FINAL REPORT.

BY- COWAN, J. N.
CORNELL UNIV., ITHACA, N.Y.

REPORT NUMBER NDEA-VI-76-1 PUB DATE 15 JAN 63
CONTRACT OEC-SAE-8365
EDRS PRICE MF-$0.09 HC-$0.76 19P.

DESCRIPTORS- *PHONETICS, *LINGUISTICS, *PHONETICS ANALYSIS, *SPEECH, LINGUISTIC PATTERNS, *PERCEPTION, ITHACA, NEW YORK

THIS 2-PART REPORT DEALT WITH THE DEVELOPMENT OF AN INSTRUMENT FOR GRAPHICALLY RECORDING SPEECH INTONATIONS. THE APPENDIXES GAVE DETAILS OF CIRCUITRY AND INSTRUMENTATION FOR THE PITCH MATCHING DEVICE AND THE INTONATION PLAYBACK. (THIS REPORT IS A SUPPLEMENT TO ED 003 881.) (LP)

ED003882

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Final Technical Report

Development of a device to record graphically intonations of speech as they are perceived by a listener. — Final Report.

U.S.O.E. Contract No. SAE-8365 July 1959-December 1961

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

J M. Cowan

Cornell University

January 15, 1963

This report consists of two parts. A general description of the method and instrumentation, together with an historical and theoretical treatment of the problem, is given as a reprint from the Proceedings of the Fourth International Congress of Phonetic Sciences which was held in Helsinki in the fall of 1961. The appendices give details of circuitry and instrumentation for the pitch matching device and the intonation playback. This latter device was under development during the period of the contract and was brought to completion after termination of the contract during the first half of 1962 by Mr. John Imre, head technician of the project. Finally, the original research proposal is appended to indicate that indeed the major objectives of the project were achieved.

Much work remains to be done in the application of these techniques to linguistic and pedagogical problems. It is hoped that other laboratories will take up part of this burden now that the basic problems of instrumentation have been solved.

APPENDIX A

Components and Circuitry of the Pitch Matching Set-up.

Components in the instrumentation shown in block diagram on page 104 of the article "Graphical Representation of Perceived Pitch in Speech", Proceedings of the Fourth International Congress of Phonetic Sciences.

Tape Loop.

This unit is a Revere model T-202 tape recorder which is used with a closed loop of tape running for about 2 1/2 seconds at 7 1/2 inches per second. Recordings are made at 7 1/2 ips and played back into the speech stretcher and physical pitch extractor at 3 3/4 ips. The internal speaker of the tape recorder is automatically disconnected when the output of the recorder feeds through an external speaker jack. The recorder runs continuously throughout the matching process.

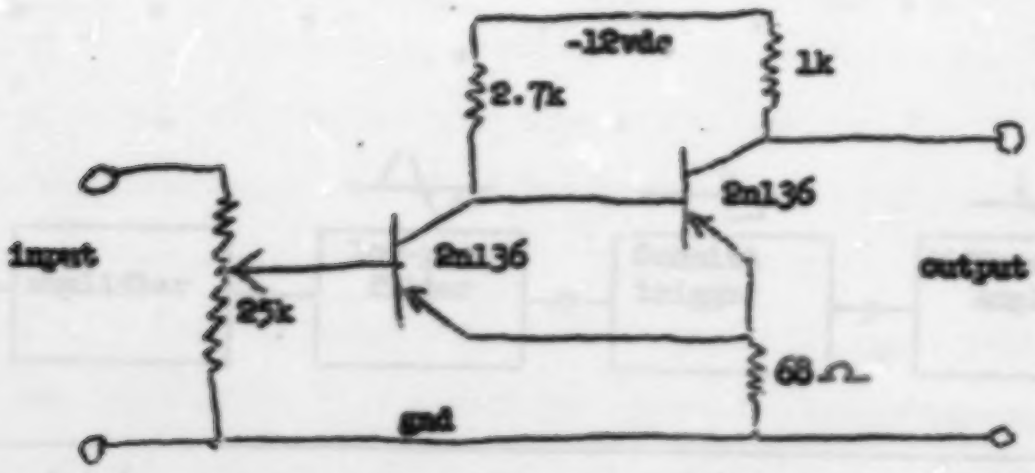
Speech Stretcher.

This unit is the standard Vari Vox manufactured by Kay Electric Company. With appropriate settings it doubles the frequencies of the input, re-recording them on a tape loop with a doubled time base.

Variable Pitch Generator.

This consists of a Hewlett Packard 207A oscillator which feeds into a Schmitt trigger. The oscillator was chosen because it is necessary to make continuous frequency sweeps between approximately 80 and 350 cps. Most oscillators have a break-point at 200 cps, at which point a multiplier switch must be used to change the scale. The circuit of the Schmitt trigger is

as follows:



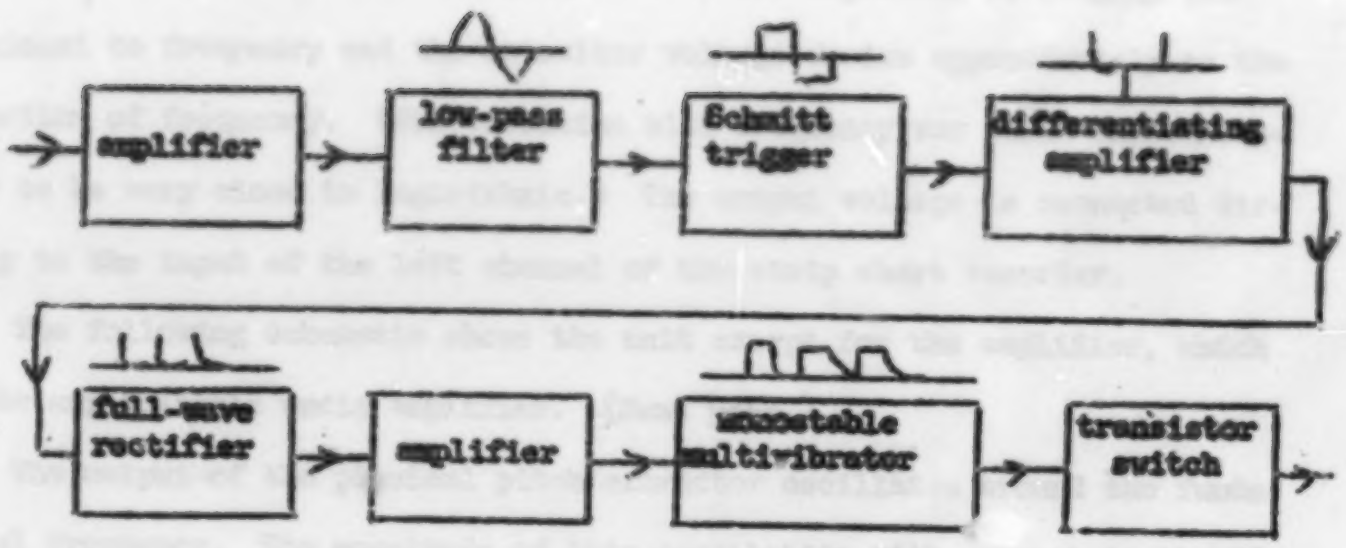
It is essentially a saturating two-stage amplifier which yields a square wave output from a sinusoidal input. The output of the trigger goes to the headphones and to a tape recorder. We used Sharpe MA-10 stereo earphones and a Nagra recorder.

Dual Channel Strip Chart Recorder.

This is a Sanborn model 320 recorder which has been modified by the addition of a solenoid which lifts the right-hand pen from the recording paper whenever the output of the variable pitch generator is manually interrupted by the operator.

Physical Pitch Extractor.

This unit may be represented by the following block diagram:



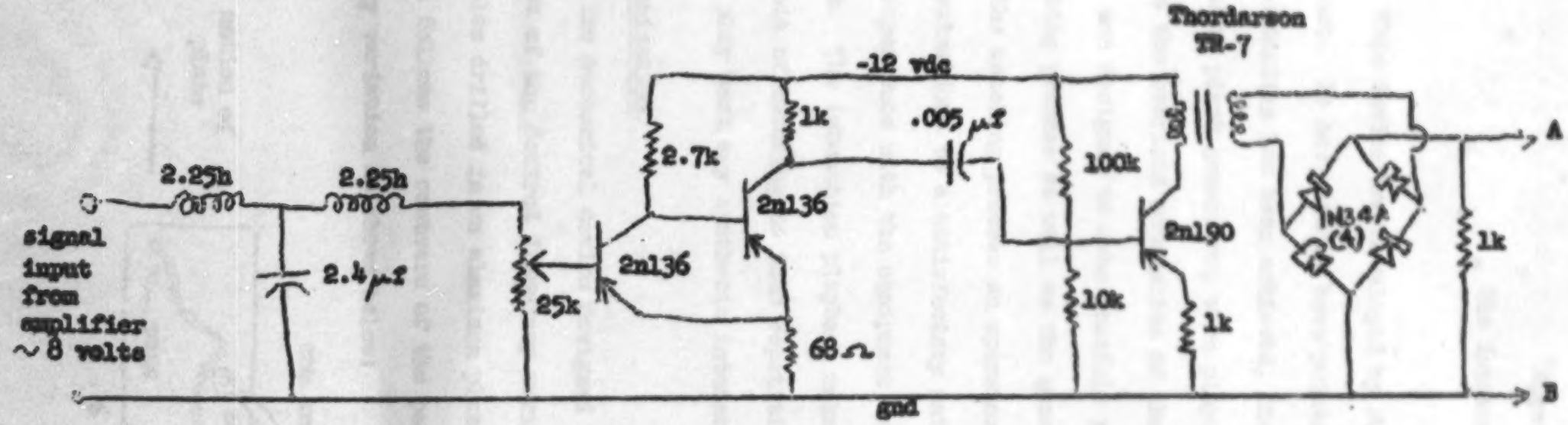
It was designed to generate an output voltage logarithmically proportional to the fundamental frequency of an input speech sample. The low-pass filter is a simple T-type designed for a break frequency of 100 cps. The Schmitt Trigger changes the passed signal into a rectilinear waveform and has the very useful characteristic of having a cutoff below which it generates no output. Since the signal passed by the low-pass filter declines above the break frequency, the Schmitt Trigger may be set by varying the input potentiometer to provide an exact cutoff at any desired frequency from 100 to slightly above 200 cps.

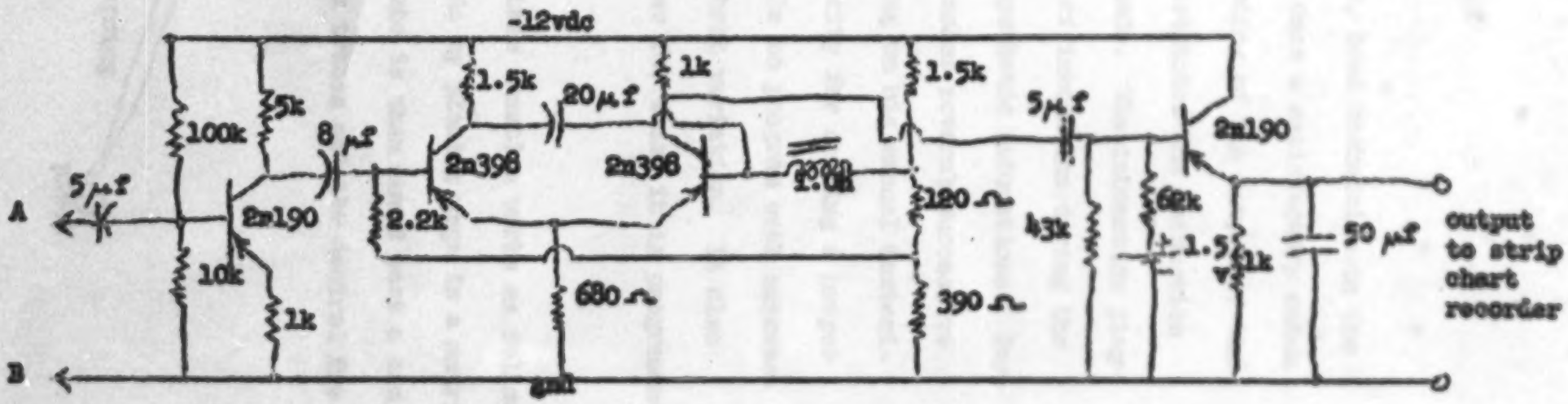
The square wave output is then differentiated and amplified. The resultant spiked waveform is then rectified by a full-wave bridge, amplified,

and fed to a monostable multivibrator. The multivibrator generates an output pulse shaped as shown. These pulses are then passed through a capacitor and are fed to the base of a transistor biased beyond cutoff. The pulses turn on the transistor for a time period roughly proportional to frequency. The emitter of the transistor is connected to ground through a 1 k resistor and a 50 μf capacitor. The charging time of the capacitor is roughly proportional to frequency and the capacitor voltage varies approximately as the logarithm of frequency. (The variation with frequency was found experimentally to be very close to logarithmic.) The output voltage is connected directly to the input of the left channel of the strip chart recorder.

The following schematic shows the unit except for the amplifier, which can be any suitable audio amplifier. (Next page.)

The output of the physical pitch extractor oscillates around the fundamental frequency. The magnitude of this oscillation with respect to the DC voltage across the capacitor at a given frequency is controlled directly by the value of capacitance chosen for the emitter capacitor. A very large value greatly reduces this oscillation, but the circuit then will not follow any but the slowest changes in pitch. Conversely a small value allows the circuit to respond faster, but introduces a greater oscillation around the fundamental. The median value of 50μf was chosen for our experiments.


Thordarson
TR-7
-12 vdc
1k
.005 µf
2.7k
100k
2n136
2n190
2.25h
2.25h
signal
input
from
amplifier
~ 8 volts
2.4µf
2n136
25k
68Ω
10k
1k
1k
IN34A
(4)
A
B
gnd


-12vdc
1.5k
20µf
1k
1.5k
5µf
2n190
100k
5k
8 µf
2n398
2n398
5µf
A
2n190
2.2k
120Ω
43k
68k
1.5
v
1k
50 µf
output
to strip
chart
recorder
10k
1k
680Ω
390Ω
B
gnd

Appendix B

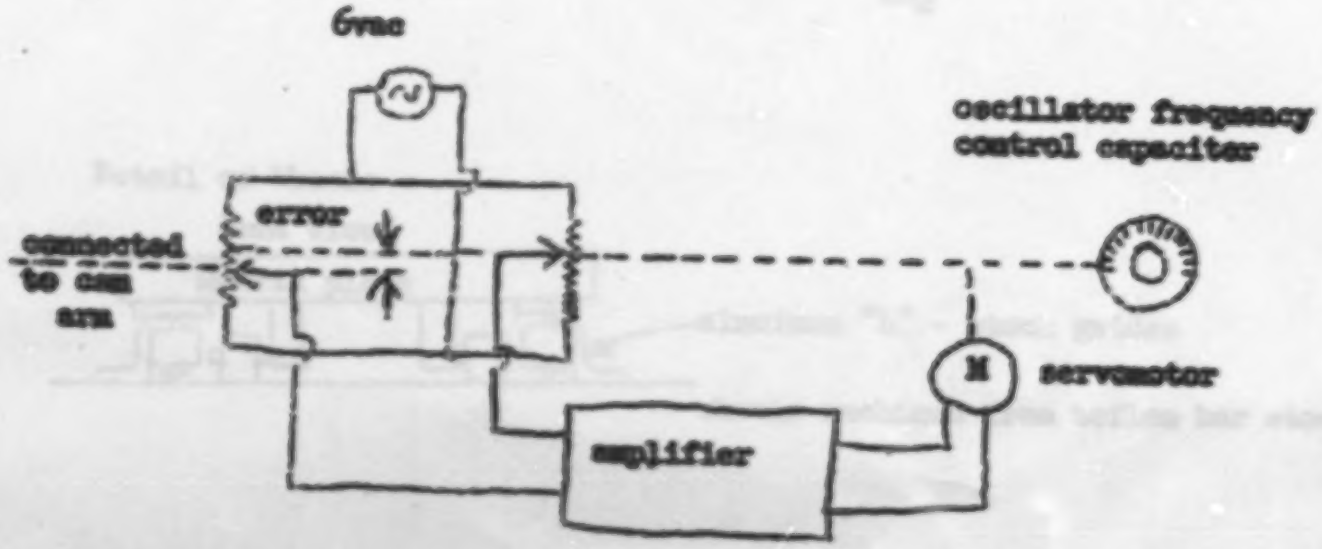
The Intonation Playback

This device was developed by Mr. John Imre, head technician on the project. It has already been pointed out that once a satisfactory match of intonation has been achieved, the tape recording of the output of the variable pitch generator, when played back, constitutes the intonation minus the overlaid modulation of the speech sounds. The intonation playback was designed to make possible programming of intonation during the matching process as well as the generation of synthetic intonations. During the matching process an operator typically makes several successive approximations to a satisfactory match, depending on his manual dexterity, experience with the equipment, and his capacity for sensing a proper match. The intonation playback makes it possible to program each successive bit of matching so that repetitions are without variation. It also will play back any synthetic intonation sequences for which it is programmed.

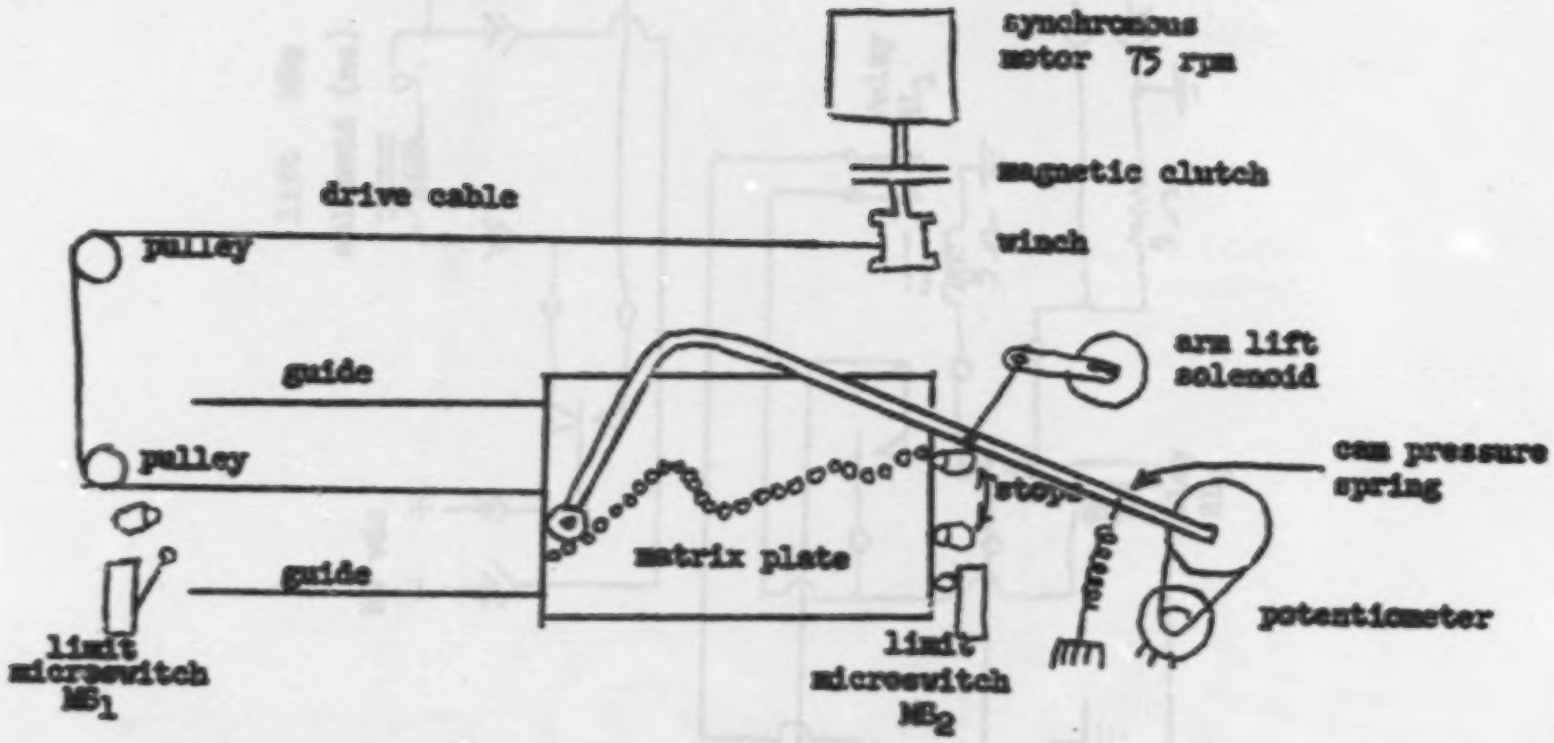
Mechanization.

The mechanical device designed to perform this function works as follows: A plot of the desired frequency variation is made by placing pegs in a matrix of holes drilled in an aluminum plate. This plate is then moved past a cam which follows the contours of the pegs. The cam traces out the desired frequency variation as shown below:

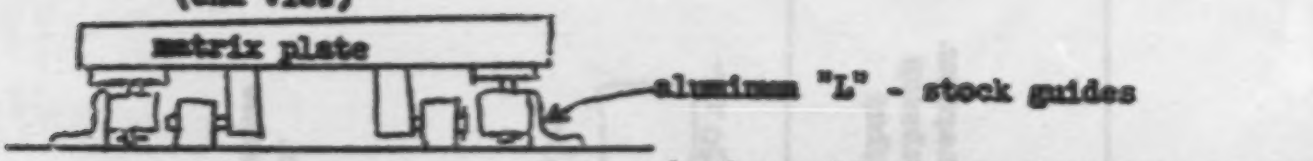
The cam arm is connected to a potentiometer so that a given angular position of the arm corresponds to a given position of the slider in the potentiometer. The outer terminals of this potentiometer are connected to those of another potentiometer which is in turn connected to the shaft of the audio oscillator. If the slider positions of the two potentiometers are not the same, there will be a difference or error voltage between their slider terminals. This error voltage is amplified and then applied to a servomotor which is connected so that it will drive the system towards the position which creates zero error voltage. This is an ordinary potentiometer position control system, and the oscillator frequency will be controlled by the location of the cam arm at any moment. The diagram of this system is shown below:



The actual physical layout and wiring schematics for this device are shown on the following pages.

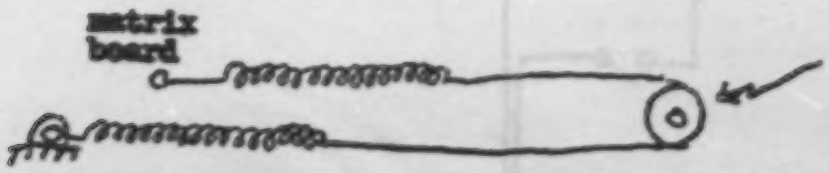


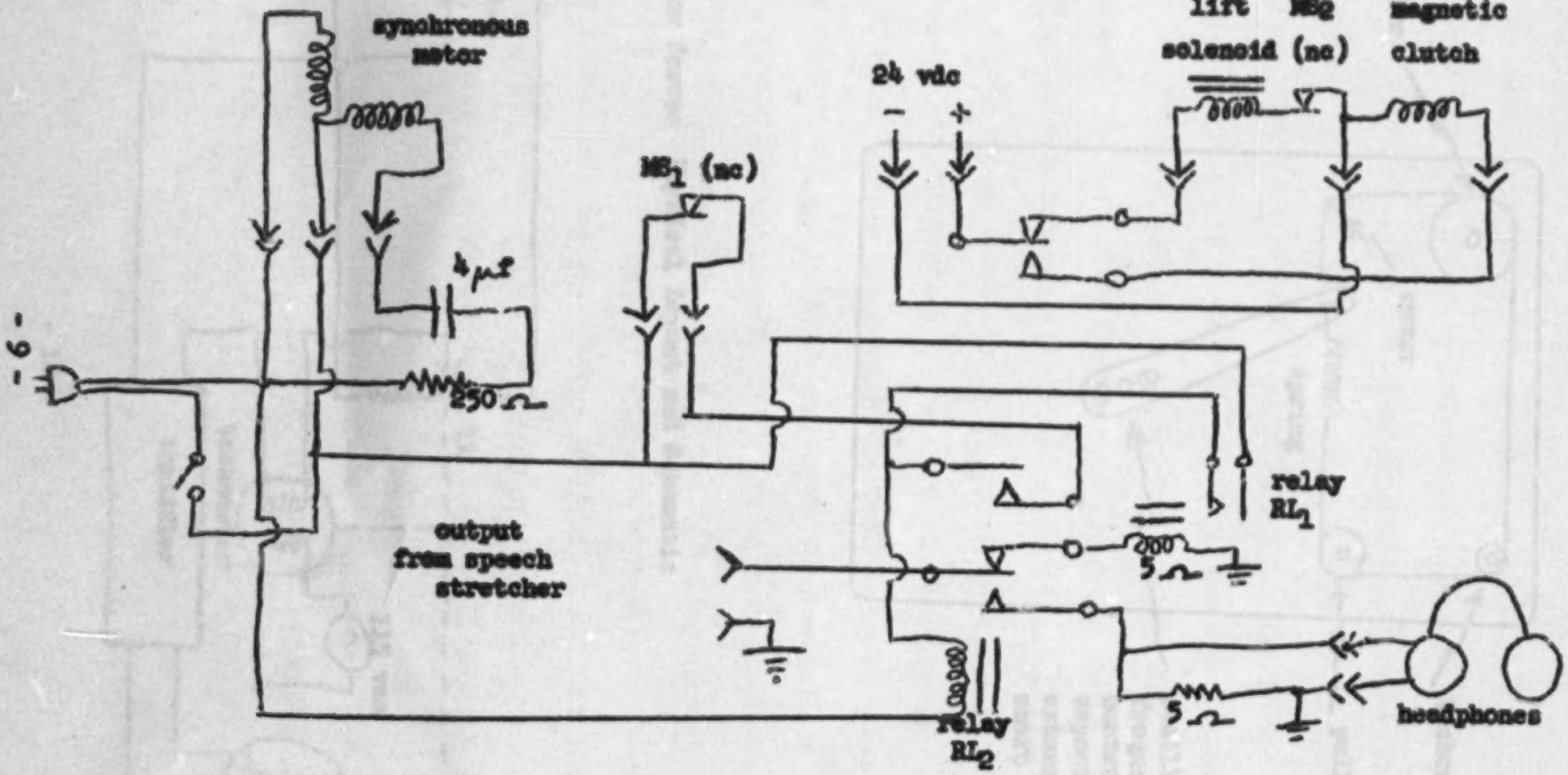
Detail of Wheels
(end view)

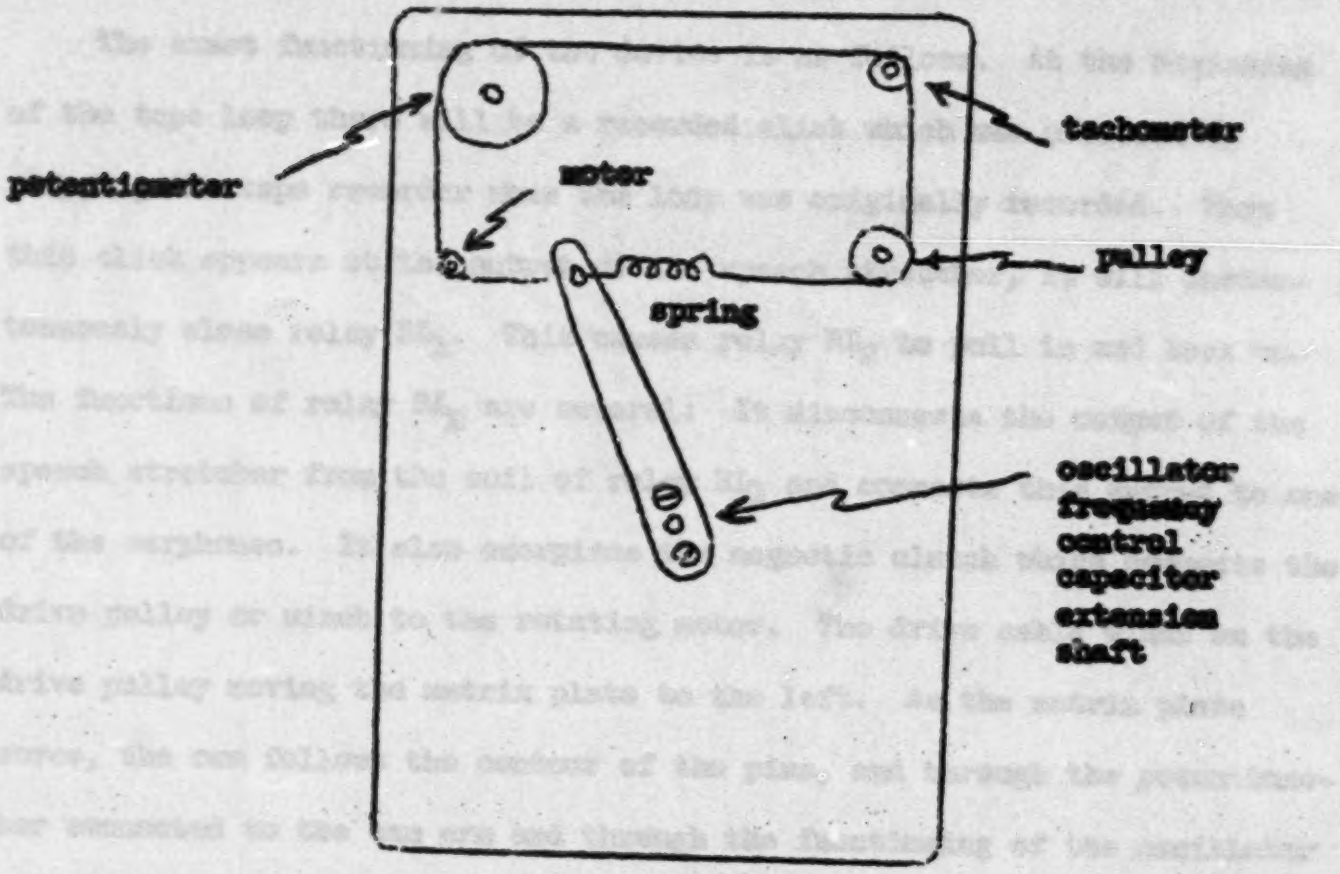


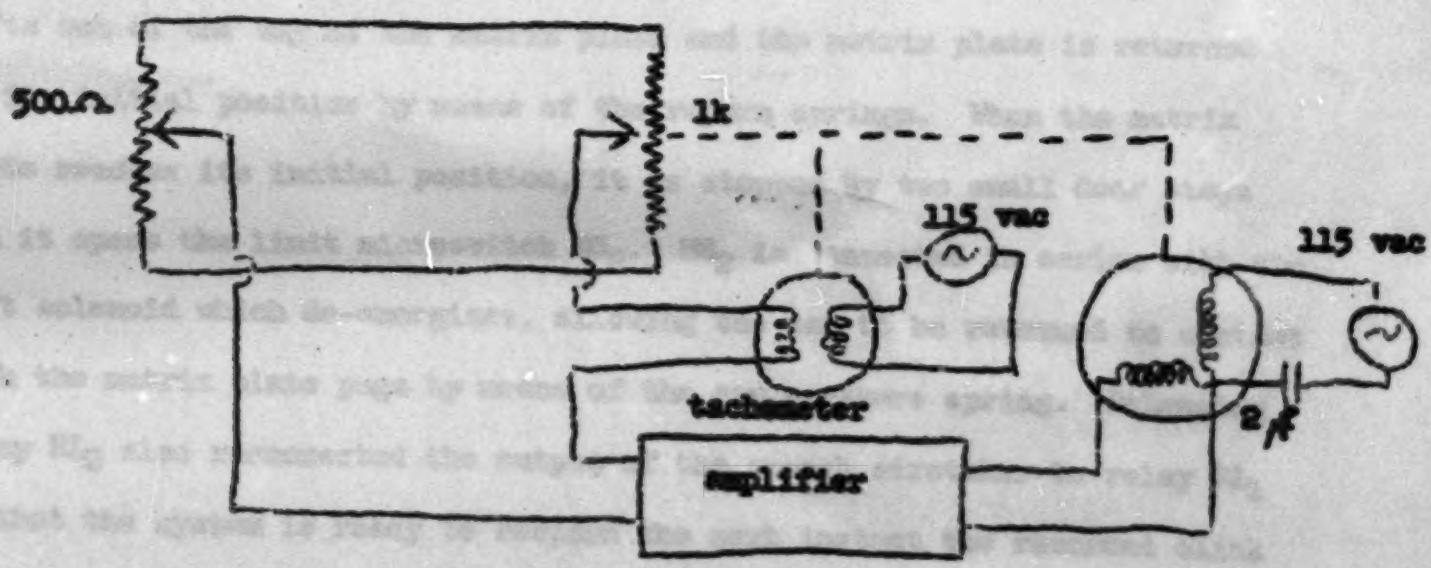
Wheels machined from teflon bar stock

Detail of Return Spring Mechanism




synchronous motor
4 μf
250 Ω
-6
MS_1 (nc)
24 vdc
–
+
lift solenoid
MS_2 (nc)
magnetic clutch
output from speech stretcher
relay RL_1
200
5 Ω
relay RL_2
5 Ω
headphones





Oscillator Servo: Physical Layout and Schematic

Functional Description

The exact functioning of the device is as follows. At the beginning of the tape loop there will be a recorded click which was produced by stopping the tape recorder when the loop was originally recorded. When this click appears at the output of the speech stretcher, it will instantaneously close relay RL_1. This causes relay RL_2 to pull in and lock on. The functions of relay RL_2 are several: It disconnects the output of the speech stretcher from the coil of relay RL_1 and connects this output to one of the earphones. It also energizes the magnetic clutch which connects the drive pulley or winch to the rotating motor. The drive cable winds on the drive pulley moving the matrix plate to the left. As the matrix plate moves, the cam follows the contour of the pins, and through the potentiometer connected to the cam arm and through the functioning of the oscillator servo, the motion of the cam produces variations in the oscillator frequency.

When the matrix plate approaches its limit of travel, it hits microswitch MS_1 which is connected in series with the locking circuit of relay RL_2. Thus the locking circuit is opened and RL_2 is de-energized. This releases the magnetic clutch and energizes the arm lift solenoid. Thus the cam arm lifts out of the way of the matrix plate and the matrix plate is returned to its initial position by means of the return springs. When the matrix plate reaches its initial position, it is stopped by two small door stops and it opens the limit microswitch MS_2. MS_2 is connected in series with the lift solenoid which de-energizes, allowing the cam to be returned to contact with the matrix plate pegs by means of the cam pressure spring. Releasing relay RL_2 also reconnected the output of the speech stretcher to relay RL_1 so that the system is ready to respond the next instant the recorded click appears in the output of the speech stretcher.

The prototype of the intonation playback at Cornell has worked well within certain limits. The relation of cam size to peg size and number of pegs limits the steepness of intonation curves which can be followed. An increase of smoothness of the cam path can be effected by weaving a thin steel tape among the pegs. There is difficulty with the oscillator servo system. The only amplifier available which had an output sufficient to drive the servo motor was not always stable for high values of gain. It would break into parasitic oscillation from time to time and would overload the servo motor. Nevertheless the device has sufficient promise to warrant further development and more careful calibration.

Appendix C

ORIGINAL RESEARCH PROPOSAL

Purpose.

To develop a device which will record graphically the intonations of speech as they are heard by a listener. A simple extension would be the development of a playback which would reproduce auditorily intonation graphs as a sequence of pitch fluctuations without the overlaid modulation of speech sounds.

Use.

Intonation is one of the most important components of a language. It is also one of the most elusive because it is overlaid with a sequence of significant phones occurring at an average rate (in most European languages) of about 12-15 per second. These latter place such a load on the attention of the listener that it is very difficult to abstract the underlying intonation from the rest of the sound complex.

A device which would analyse out the intonation and represent it graphically (and perhaps reproduce it auditorily as sheer intonation) would give the linguist a new and reliable tool for the analysis and study of this important component. Such a device would have important application in the teaching of foreign languages because it would illustrate vividly to learners the differences between their native intonation patterns and those of the target language. It is common knowledge that these native patterns are so deeply ingrained (before the age of six) that a learner of a foreign language may approximate very closely the imitation of the individual phones of the foreign language and still speak it almost unintelligibly or with a very heavy accent simply because of the carry-over of his own intonation patterns. Witness the way in which most Frenchmen speak English!

History.

Because of its importance, intonation has in the past been much studied with meager results. Two main lines of research have been undertaken. One is strictly impressionistic. A listener concentrates on intonation and attempts to record the pitch fluctuations he hears on some sort of arbitrary scale. No two listeners have ever yet been calibrated so they can come up with identical results for the same sample of speech, even if the speech is recorded and available for multiple playback.

The other approach attempts to give a graphical representation of the progression of the vocal cord tone. The most extensive work along this line has been done by Cowan (see attached score). He pointed out that the intonation curves which he produced by a very time-consuming measurement of oscillograms did not represent the intonations which a listener perceived when listening to the speech from which the curves were derived. In short, perceived (heard) pitch is not uniquely determined by physical (frequency in c.p.s.) pitch. Other investigators have improved the techniques of graphical recording of the cord tone progression but these advances have not gotten around the fact that it is the human ear-brain which determines what will be heard (perceived) in the way of intonation in language.

Procedure.

A device which will graph or reproduce intonation must be a hybrid of electronic hardware and a human ear-brain. Cowan has been experimenting for years with such hybrids. Basically the operator listens to speech through earphones and imitates the perceived intonation manually through controls which guide a variable tone generator. The original signal (speech) is fed

in one ear and the synthetic tone is fed in the other ear. The operator repeats the synthetic imitation of any speech sample until mismatches between the two ears are eliminated at which time the imitation intonation equals the perceived intonation. The generator-synthesizer graphs continuously on a logarithmic scale. Consequently when the synthetic signal matches the speech input, the graph represents the perceived intonation. It is a simple matter to arrange to have such a graph operate the controls of the variable tone generator so that the (matched) imitated intonation will be played back audibly.

It goes without saying that the operator who is the human component in the hybrid must undergo considerable training. However, experience has shown that different operators working over the same language sample come up with identical intonation scores which removes all doubt about the validity of the procedure.